

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: May 29, 2002, 11:38:48 ; Search time 34.14 Seconds
(without alignments)
3278.492 Million cell updates/sec

Title: US-08-485-355b-50
Perfect score: 3374
Sequence: 1 MGDAGVASQRPNNRGTNRV.....GKIAARVARRARRAARAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3374	100.0	647	12	082462 helicoverpa
2	2205	65.4	644	12	090063 nudaurelia
3	435.5	12.9	757	12	09YK87 thossea asig
4	386.5	11.5	612	12	09YR82 nudaurelia
5	134	4.0	2147	2	09L950 pseudomonas
6	132	3.9	886	12	09Q887 human herpe
7	131.5	3.9	856	12	056854 human herpe
8	130	3.9	597	16	09RY11 deinococcus
9	127	3.8	466	5	09S510 drosophila
10	127	3.8	1212	5	09ST60 drosophila
11	127	3.8	2016	5	09V439 drosophila
12	127	3.8	2016	5	09NBA1 drosophila
13	126	3.7	3570	4	099552 homo sapien
14	125.5	3.7	717	2	09KHB3 bacillus ci
15	125.5	3.7	1269	17	097206 sulfotolopus
16	125	3.7	2314	2	069822 streptomyc

17	123.5	3.7	1204	17	09UY51 pyrococcus
18	122	3.6	1314	4	09UMZ2 homo sapien
19	121.5	3.6	699	2	048494 kurtzia zop
20	120.5	3.6	769	12	09P2T4 human parvo
21	120.5	3.6	781	12	09J318 human parvo
22	120.5	3.6	979	2	09XAS7 streptococ
23	120	3.6	873	16	09RX36 delnoco
24	119.5	3.5	851	12	09OP86 human herpe
25	119.5	3.5	889	4	014917 homo sapien
26	119.5	3.5	907	12	066537 human herpe
27	119.5	3.5	1425	5	09VIT9 drosophila
28	119.5	3.5	2271	16	099QY4 staphylococ
29	119	3.5	1993	16	092RG5 listeria in
30	119	3.5	2809	5	061230 lytechinus
31	118.5	3.5	3029	16	055582 syncocyst
32	117.5	3.5	781	12	09UGP8 human parvo
33	117.5	3.5	884	16	09RMT0 delnoco
34	117.5	3.5	26926	4	010466 mycobacteri
35	116	3.4	4578	16	050379 fuqu rubrip
36	116	3.4	4578	13	042181 mycobacteri
37	115.5	3.4	1496	5	09NFV5 drosophila
38	114.5	3.4	781	12	065789 human parvo
39	114.5	3.4	828	10	089316 human parvo
40	114	3.4	828	10	004238 victia faba
41	113.5	3.4	658	12	066536 human herpe
42	113.5	3.4	781	12	085117 human parvo
43	113.5	3.4	781	12	090200 human parvo
44	113.5	3.4	781	12	090221 human parvo
45	113.5	3.4	781	12	090222 human parvo

ALIGNMENTS

RESULT 1
ID 082462 PRELIMINARY: PRT: 647 AA.
AC 082462:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COAT PROTEIN.
GN P71.
OS Helicoverpa armigera stunt virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage: Tetraviridae;
OC Unclassified Tetraviridae.
OX NCBI_TaxID=37206;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BLACK MOUNTAIN;
RX MEDLINE=97201539; PubMed=9049325;
RA Hanzlik T.N., Dorian S.J., Johnson K.N., Brooks E.M., Gordon K.H.;
RT "Sequence of RNA2 of the Helicoverpa armigera stunt virus
(Tetraviridae) and bacterial expression of its genes.";
RL J. Gen. Virol. 76:799-811(1995).
DR EMBL: L37299; AAC37885.1; -;
DR MEROPS: A21.001; -;
SQ SEQUENCE 647 AA: 70670 MW: 08205B8CB53CFBD2 CRC64;

Query Match 100.0%; Score 3374; DB 12; Length 647;
Best Local Similarity 100.0%; Pred. No. 1e-230;
Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDAGVASQRPNNRGTNRVSVANTVYVNGRRNORRTGROVSPDNFTAAADLAOSL 60
DB 1 MGDAGVASQRPNNRGTNRVSVANTVYVNGRRNORRTGROVSPDNFTAAADLAOSL 60
QY 61 DANTYFPANISSMPEFNMAKGRIDLDSDSGVFXFLDPAGATESARAAGEVSKIIPDG 120
DB 61 DANTYFPANISSMPEFNMAKGRIDLDSDSGVFXFLDPAGATESARAAGEVSKIIPDG 120
QY 121 LVKFSVDAEIRIYNECPVVTDVSVPLDGRWLSLIFSFPNFTAVVAANVENKMSL 180

```
Db 121 LVFSDAEIREIYNECPVYTDVSVPLDGRQWSLSIFSFPMEFRTAVVAANENKMSL 180
Qy 181 DVNDLLEIMNLADWRYVDSSEOMINFETNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 240
Db 181 DVNDLLEIMNLADWRYVDSSEOMINFETNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 240
Qy 241 YRLTYKAITCEANMPTLVDOGFWMIGGOYALPTSLPOYDVESEAYALHTLTFARPSSAAL 300
Db 241 YRLTYKAITCEANMPTLVDOGFWMIGGOYALPTSLPOYDVESEAYALHTLTFARPSSAAL 300
Qy 301 AFWAGIIPGCGTAPAGTPAMEQASSGGYLTWRHNGTTFPAGSVSYLPEGFALERDND 360
Db 301 AFWAGIIPGCGTAPAGTPAMEQASSGGYLTWRHNGTTFPAGSVSYLPEGFALERDND 360
Qy 361 GSWTDFASAGDTVFRQVADEVYVYTNPNPAGGSAPFTYRVPSPNAYTNTVFNLTLET 420
Db 361 GSWTDFASAGDTVFRQVADEVYVYTNPNPAGGSAPFTYRVPSPNAYTNTVFNLTLET 420
Qy 421 RPSRRLELPMPPADFGQTVANNPKIEOSLLEKTLGCVLHVKMRNPFOLTPASSFGAV 480
Db 421 RPSRRLELPMPPADFGQTVANNPKIEOSLLEKTLGCVLHVKMRNPFOLTPASSFGAV 480
Qy 481 SFNNPGYERTRLPDYTGIRDSFDQNNSTAVAHFRSLSHSCSIYTKYQGWEGVTNNP 540
Db 481 SFNNPGYERTRLPDYTGIRDSFDQNNSTAVAHFRSLSHSCSIYTKYQGWEGVTNNP 540
Qy 541 FGGFAHAGLKNELICLADDLATRLTGVPATDNFAAASAFANNLSSVLKSEATSI 600
Db 541 FGGFAHAGLKNELICLADDLATRLTGVPATDNFAAASAFANNLSSVLKSEATSI 600
Qy 601 IKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
Db 601 IKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
```

```
RESULT 2
Q90063 2 PRELIMINARY: PRT: 644 AA.
AC Q90063;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE CAPSID PROTEIN.
GN CAPSID PROTEIN. CP.
OS Nudaurelia capensis Omega virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC Omegatetravirus.
OX NCB1_Taxid=12541;
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92391097; PubMed=1519360;
RA Agrawal D.K., Johnson J.E.;
RT "Sequence and analysis of the capsid protein of Nudaurelia capensis
RT Omega virus: an insect virus with T = 4 icosahedral symmetry.";
RL Virology 150:806-814(1992).
DR EMBL: S43937; AAB23198.1; -.
DR MEROPS: A21.001; -.
SQ SEQUENCE 644 AA: 69863 MW: EAA1C9AB50F24828 CRC64;
```

Query Match 65.4%; Score 2205; DB 12; Length 644;
Best Local Similarity 65.8%; Pred. No. 7.1e-148;
Matches 427; Conservative 69; Mismatches 135; Indels 18; Gaps 6;

```
Qy 8 SQRPHRRGRNRYVSNANTYVNGRRNOR--RRTGROVSPDNFTAAADLQSLDANTY 65
Db 3 SNSASKRSRNRVRIANTYVNAVKORQARGRRARRANNIDVTAADLQSLDANTY 62
Qy 66 TEPANISMPPEFRNMAKGIKIDLDSDSIGWFKYLDPAGETESARAAGEVSKIPDGLVKS 125
Db 63 TTFNTVATMEPEFRNMAKGIKIDLDSDSIGWFKYLDPAGETESARAAGEVSKIPDGLVKS 122
```

```
Qy 126 VDAEIREIYNECPVYTDVSVPLDGRQWSLSIFSFPMEFRTAVVAANENKMSL 185
Db 123 VDAEIREIYNECPVYTDVSVPLDGRQWSLSIFSFPMEFRTAVVAANENKMSL 182
Qy 186 LIEMNLADWRYVDSSEOMINFETNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 245
Db 183 LIEMNLADWRYVDSSEOMINFETNDTTYVYRIRLRLRPYDVPDEGLVPTVSD 242
Qy 246 KAITCEANMPTLVDOGFWMIGGOYALPTSLPOYDVESEAYALHTLTFARPSSAALAFVNA 305
Db 243 KAITCEANMPTLVDOGFWMIGGOYALPTSLPOYDVESEAYALHTLTFARPSSAALAFVNA 302
Qy 306 GLPQGGTAPAGTPAMEQASSGGYLTWRHNGTTFPAGSVSYLPEGFALERDNDSDWD 365
Db 303 SMPGGSAPGSDPAMIIPDSTGTC-QWRHGGFDAPGTIVITIPRGYTMQVFTTNNEMNG 361
Qy 366 FASAGDTVFRQVADEVYVYTNPNPAGG-----SAPFTYRV-PPSNATNTNY-FNLTLL 418
Db 362 FANPDVYTFGQ-----TGGAAGTNTITTTAFTVTLTTLATTTSAANVINFRNLDA 413
Qy 419 RPSRRLELPMPPADFGQTVANNPKIEOSLLEKTLGCVLHVKMRNPFOLTPASSFG 478
Db 414 ETTAASNRSEVPLPLFGOTAPANNPKIEOTLVKDTLGSLVHVKMRNPFOLTPASSFG 473
Qy 479 AVSFNNPGYERTRLPDYTGIRDSFDQNNSTAVAHFRSLSHSCSIYTKYQGWEGVTNN 538
Db 474 AISFTNPGFDRNLDPFGFGRISLDVNNSTAVCHFRSLSHSCSIYTKYQGWEGVTNN 533
Qy 539 TPRGOFHAGLKNELICLADDLATRLTGVPATDNFAAASAFANNLSSVLKSEATS 598
Db 534 TPRGOFHAGLKNELICLADDLATRLTGVPATDNFAAASAFANNLSSVLKSEATS 593
Qy 599 SIKSVGETAVGAOSGLAKLPGLIMSVPGKTAARVARRARRARRARAN 647
Db 594 SVIKELGNQATGLANQGLARLPGLASIPGKTAARVARRARRARRARAN 642
```

```
RESULT 3
Q9YK87 3 PRELIMINARY: PRT: 757 AA.
AC Q9YK87;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE CAPSID PROTEIN.
GN TAY-CP.
OS Thosea asigna virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC unclassified Tetraviridae.
OX NCB1_Taxid=83810;
RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99350019; PubMed=10423156;
RA Pringle F.M., Gordon K.H., Hanzlik T.N., Kalmakoff J., Scotti P.D.,
RA Ward V.K.;
RT "A novel capsid expression strategy for Thosea asigna virus
RT (Tetraviridae).";
RL J. Gen. Virol. 80:1855-1863(1999).
DR EMBL: AF062037; AAC97195.1; -.
DR MEROPS: A21.UNM; -.
FT CHAIN 1 155 UNKNOWN.
FT CHAIN 156 695 LARGE CAPSID PROTEIN.
FT CHAIN 696 757 SMALL CAPSID PROTEIN.
SQ SEQUENCE 757 AA: 82413 MW: 17432C70D2C427C CRC64;
```

Query Match 12.9%; Score 435.5; DB 12; Length 757;
Best Local Similarity 25.9%; Pred. No. 2.4e-22;
Matches 170; Conservative 105; Mismatches 263; Indels 119; Gaps 27;

```
Qy 26 TYTVNGRRNORR---TGROVSPDNFTAAADLQSLDANTYTPPANISMPPEFRNMAK 82
Db 174 TKTAPGKRTTRRRRRNGKGTGNPNPFSISA-----GAPMSKIPALRGTLA 219
```

Query Match	11.5%;	Score 386.5;	DB 12;	Length 612;
Best Local Similarity	26.0%;	Pred. No. 5.2e-19;		
Matches 174;	Conservative 86;	Mismatches 295;	Indels 115;	Gaps 31;

QY	15	KCTNVRVRSANTVYVYNGRRNRORRTGROVSP----	PDNFATAAODLAOSLDANTVTFPANI	71			
Db	8	RPARNPSSOGNGRRNNNNKRRRRRRGLKLPVVA*-	ITSGQMAEPANHANT-----	58			
QY	72	SSMPFEFRMAKQ-----	KIDDSISLIMFKYLDPAATESARAVGYSKIPGLKFS	122			
Db	59	-RVNGRTVRVRLQDAMMESPAATASEAMHIDYLD	PDDEYTSLDDQ--KIPGALPOS	111			
QY	136	VDAEIREIYNEBCPVVTVSVDPLDGRQMSLSIFSE	PFMRTAIVVAVANENKMSLDVVN-	184			
Db	115	TCGGRGVGVARGVAPGLNSTPTLPLDGGTMBLMLH	PFRRHLLFTTSSNE--VEYVNA	177			
QY	185	DL-----TEKLNLDAMRWYVDSEQMINFTND	TTYVIRLVRLPRT---YVDPPPE-GLVR	233			
Db	173	DLDAEANDM-NKRTWTEAT-YPSAAQVGN--	VEYMAV----PREALDVPPTPOLGVSG	224			
QY	237	TVSDRLVFKATTCANPTELDVQGFWMJGGQALP	TPSLPODYSEVALHTLT-----	290			
Db	225	LLESTRLLSSGTALFAPNAPTLVNOGVAVIAQF--	QPKENQKENPDVAGTTQIGGTLQL	287			
QY	291	-FARSSAALALFVWAGLPQGSTA-PAQTPAMEQAS	SGGYLTWRHNGTTPRAG--SVSY	344			
Db	283	GGSGBNNTLTMT-IGDOYEFGGALPLPTVSKG	PMPESGQLVEQTANLTFDVGMTITTT	341			
QY	346	VLPREGALERDPNNGSMTDFASAG-DVTFPQVAV	DEVVYTNPNAGGSGAPTFVVRPP	400			
Db	342	TLPPGSV-----TGMOQTASNGTDIVT-----	VD-----ACARLYAGCANLDA	380			
QY	405	SAAYNTVEFRNTLTETPSSRLRLPMPADGCGT	VANNPKIEGSLKETLGCYLVHSM	466			
Db	381	SELNIQDI-----NSKIKIPTNNQMQMTPTT	IQQLNETGFLTH-PLRA	422			
QY	465	RNPVEQLTPASSFGAVSFNPNGETRDLDPD----	TGIRSDFDQNMSTAVAHFRSLSHS	520			
Db	426	FQPVEMTMTATSGYGVWKTP---RT-TVVDHRAI	GGIDQTDISNFAIGVAAAMTGMS	483			
QY	521	CSIVTKTQGWGCVNVNMPRGQAFNAALLNKEEL	CLADDLATRLTGVPATDNFAAV	580			
Db	482	TVYFVYKVRFRFALPADESSPMGPFASAPRP	DDVALVARTWTDLHPAYERYNGEGL	543			
QY	581	SAFANMLSVYKSEATSSIIKSVETAVGAOSGLA	LPGLMSVPKIAARYARARAR	640			
Db	542	FAMVAKTIAQIPR-----YVRS-----	AAGVANAATDIEATESVASNSISRRQRAR	591			
QY	641	R-----RAAR	645				
Db	592	RVGGIARGAR	601				
RESULT	5						
Q9L950							
AC	Q9L950	PRELIMINARY;	PRT; 2147 AA.				
DT	01-OCT-2000	(Tremblrel. 15, created)					
DT	01-OCT-2000	(Tremblrel. 15, last sequence update)					
DT	01-DEC-2001	(Tremblrel. 19, last annotation update)					
DE	OUTER MEMBRANE-LIKE PROTEIN.						
OS	Pseudomonas putida.						
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;						
CC	Pseudomonas.						
OX	NCBI_TaxID=303;						
RN	[1]	SEQUENCE FROM N.A.					
RP	MEDLINE=20225830;	PubMed=10762233;					
RA	Espinosa-Urgel M., Salido A., Ramos J.L.;						
RT	"Genetic analysis of functions involved in adhesion of pseudomonas						
RT	putida to seeds."						
RL	J. Bacteriol. 182:2363-2369(2000).						
RP	[2]	SEQUENCE FROM N.A.					
RP	Espinosa-Urgel M., Salido A., Ramos J.L.;						
RL	Submitted (May-2000) to the EMBL/GenBank/DBJ databases.						

DR EMBL, AF182515; AAF69025.1; -
DR InterPro: IPR001343; HemIysn_Ca_bind.
DR PRINTS: PRO013; CABNDGRPT.
SQ SEQUENCE 2147 AA; 211587 MW; BD7EC3B46FE2D6B CRC64;

Query Match 4.08; Score 134; DB 2; Length 2147;
Best Local Similarity 22.08; Pred. No. 2.8;
Matches 141; Conservative 75; Mismatches 264; Indels 162; Gaps 33;

QY 39 TGRVSPDP-NFTAAADLAQSLDANTVTFPANTSSMPEFRNNAKGRIDDSI-----GW 94
DB 1363 TGTALADLTTPPAAGLFLFSADGCTLTGEGATVTVS--AGTV-LQTATVQADGT 1420
QY 95 YFKYLDAGATASARAVGESKIPDGLVKSVDAREIYNEECPPVTD-----VSPLD 149
DB 1421 FTVTLSP--PDQNGQLLSVLTSDPRGNVSGVNVYTPADV-DANAPVLAQDLATATVLA 1477
QY 150 G-----RQMSLSTSFPMFRRTAYAVANVENKESLDVYNDLIEMNLADMRVVD 201
DB 1478 AVSATKTYTDSFTLLSGFSKTYMTVAAGTTADPTLLT--TNSVLAALNNTSFTLQVKD 1536
QY 202 -SEQMINETNDTYVYRIRVRYDYVDPDEGL-----VTVSDYRLTYKA--ITCEAN 253
DB 1337 ASGANVITATGNTQGLDLVLP-----GLQVDIGVLQAGDRLTVGSGIGLITE 1588
QY 254 MPTLVDFGFMIGQYALTPSLPY-----DVSEAYALHTLTFAR--PSSAA 299
DB 1589 VSTLLD-----IVSTSLTQFTGTGATSGNVITDVGTCDAVD---ARCPDSAAV 1634
QY 300 LAFWAGLIPQCGTAPACTPAMEQASSG-----GYLWRKNGTTFPAGSV--SYVL-- 347
DB 1635 LQV-----LKDSYVSAGTATTVQOQYGTLVIRADGSYTYTPNGSPNSGKDVESYOLVH 1650
QY 348 PEGR--ALERY-----DPNDGSWTDFASAGDTTFEQVAVD--EVVYNNPAG 391
DB 1691 PNGLSDAANLYVRIDSPQATFVMSDITNGAPATVYDAVNDVGTQVYLLDRNVDSSSTLG 1750
QY 392 GGSAPFTTVRPPSNNAVTVFNKTLLETPRSSRLELPMRPADGOTVANNPKIEQSL 451
DB 1751 SLNLPVIGTR---QATYTTTVAANTADLQVNVSTNL-----LSLNGTTEIELKL 1799
QY 452 KETLGCYLVHSKMRNPFOULPRASSFGAVSFNNNGYERTRLPYTGIRSFQDMSTAV 511
DB 1800 NPATGQTV-----LVQSVPGSLVSLGGAGY-----TF-ENQAGT 1836
QY 512 AHFSLSHSCSIYTKYQEGEVTVNTP--FGQFAHAGLKNELICLADLADLRLTG 568
DB 1837 YHNVATAGGIGLSSI-----TTSLNITTTTYLLEFYVSG-----ATAATG 1876
QY 569 VVPATDNFAAVSAFANMLSSVLKSEATSSIIKSVGETAVG 610
DB 1877 NLADDMVGSALTIVL-----SVLTAAVTYTIIPGNGVSAG 1912

RESULT 6
Q90P87 PRELIMINARY; PRT; 886 AA.
AC Q90P87;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER ENVELOPE GLYCOPROTEIN GP350.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98203772; PubMed=9542645;
RA Lee W.K., Kim S.M., Sim Y.S., Cho S.G., Park S.H., Kim C.W.,
RA Park J.G.;
RT "B-lymphoblastoid cell lines from cancer patients.";

RL In vitro Cell. Dev. Biol. Anim. 34:97-100(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99072166; PubMed=9856346;
RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,
RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;
RT Cloning and analysis of the Epstein-Barr virus glycoprotein 350
genes.";
RL Mol. Cells 8:585-593(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,
RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF172332; AAD51697.1; -
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
KM Envelope protein.
SQ SEQUENCE 886 AA; 92487 MW; F4BCA573EBE29A48 CRC64;

Query Match 3.9%; Score 132; DB 12; Length 886;
Best Local Similarity 20.5%; Pred. No. 1;
Matches 139; Conservative 66; Mismatches 268; Indels 206; Gaps 29;

QY 48 NFTAADLAQSLDANTVTFPANTSSMPEFRNNAKGRIDDSIGWYFYLDPAAGATES 107
DB 169 NITAVNR--AQGLD---VTLPLSLTPSAQDSNFS-----VKTELG--NEID---IEC 211
QY 108 ARAVGEYSKIPDGLVKSVDARE-----IYNECPVTVDSVPLDQRQMSLSIFR 160
DB 212 IMEDGELISQVLPQGNKNITVCSGESHVPSGGLTSTSPAT--PLPTGYAASLRTPR 269
QY 161 PMFR-----TAYVAVANVENKESLD-----VVNDLI-----EWLNADMRVYDSE 203
DB 270 PVSRLGNNSILYFYSGNGPKASGSDYCIQSNIVFEDELPASODMPTNTDITLYGDN- 328
QY 204 QWINTFTDTTYVYRIRVLRPTVYDPDEGLVRTVSDYRLTYAICEANMPLVDGFW 263
DB 329 -----ATYSVPMTSE-----DANSPVMTVAFW 352
QY 264 -----IGQYALMPT----- 273
DB 353 AMPNNETDKCKWTLTSGTPSCGENTSGAFASNRFTDITVSGLGTAPKTLITTRATNA 412
QY 274 -----SLPODYSEAYALHTLTFARSSAAALAFVWAGLPQGGTAPAGTPAMEQAS 324
DB 413 TTTTHKVIYSKAPESSTTSPTLMTTGFAAPNT-----TGLPSSTHVPTNLTA--PTIS 463
QY 325 SGGILWRHNGTTFPAGSVYLPBGFALERYDPNDG-----SWTDFASAGDYTFRQ 377
DB 464 TGPIVSTADYTSPTPAQTSGASP--VTPRPSPRDNGTESKAPDMTSPTSAAVTTPPNA 520
QY 378 VAQDEVVVT-----NNPAGGSAPTFTVRVPPNAVNT-----VERNNLLEFRPS 423
DB 521 TSETPAVTTPTPNATSPTLTKTSPTSAVTTPTPNATSPTPAVTTPPNATITPLTGTSPT 560
QY 424 SRRLLEPMRPADGQTVANNPKIEOSLKEKTCYLVHSKMRNPVQLTPASSFGAVSFN 483
DB 581 S-AVTPPTPNATSPYGETSPKANNT--NHITLG-----GTSSTPVYTSPPKNTSAVTIG 632
QY 484 NPGYERTRLPDYTGIR-DSFDQNMSTAVAHFPSLSHSCSIYTKYQEGEVTVNTPPG 542
DB 633 Q--HNITSSSTSMISRPSSISSETLSPTSD-NSMSHMLPLTSAHPTGCGENTIQV--TPAS 688
QY 543 QFAHAGLKNELICLADLADLRLTGVPAT-----DNEFAAFAFANMLSSVLKSE 555
DB 689 TSTH-----HVTSSPAPRPCTTSQASGPGNSSSTRKGEVNVTKGTPPKN 734
QY 596 ATSSIIKSVGETAVGAOS 614
DB 735 ATSPQAPSGGKTAVPYVTS 753

RESULT 7
ID 056854 PRELIMINARY: PRT: 856 AA.
AC 056854:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GP340.
GN BLRF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
RX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M81;
RX MEDLINE=97082049; PubMed=8923292;
RA Mackelt M., Cox C., Pepper S.D.V., Lees J.F., Beverley B.A.,
RA Wedderburn N., Arrand J.R.;
RT "Immunization of common marmosets with vaccinia virus expressing
RT Epstein-Barr virus (EBV) gp340 and challenge with EBV."
RL J. Med. Virol. 50:263-271(1996).
DR EMBL: X99106; CA67558.1;
SQ SEQUENCE 856 AA; 89634 MW; E34E581AF1071CDD CRC64;

Query Match 3.9%; Score 131.5; DB 12; Length 856;
Best Local Similarity 20.8%; Pred. No. 1;
Matches 120; Conservative 51; Mismatches 190; Indels 217; Gaps 23;

QY 48 NPTAAADLAQSLDANTVTFPANISSMPEFRNMA-----KKRIDSDSIGMYEKYLD 101
DB 169 NITAVVR--AQGLD--VTLPLSLPTSAQDSNFSVKTMGLNGEID----- 209
QY 102 ACATSAARAVGEYSKIPDGLVFSVDAIRE-----IYNECPYVDVSVPLDGRWS 154
DB 210 ----ECIMEDGEISOVLPDKNFNITSCGESHVPSGGILITSPVAT--PIPGTGYAS 263
QY 155 LSIFFSPMR-----TAVVAANVENKEMSLD-----VYNDLI-----EMLNLMADR 197
DB 264 LILTRPVRFRIGNSILYFSGNPKRSGDYCIQSNIIVSDELIPASQDMPTTTDT 323
QY 198 YVVDSEQWINTNDTYYVIRIVLPRTYDVPDTEGLVTVSDYRLTYKAITCEANMPTL 257
DB 324 YGVDN-----ATYSVPWVTSE-----DANSFV 346
QY 258 VVOGF-----IGGOYALTP----- 273
DB 347 TVTAFWAMENNETDEKCKMTLTSCTPSCENISGAFANRFDITVSGLTAPKTLIT 406
QY 274 -----SLPDYVSEAYALHTLFARPSAAALAFWAGLPQGTAPAGT- 317
DB 407 KRAITNATTTTHKVIKSKAPESTTTSTPLTWTGTFADPNTTGLPSS--THVPTMLTAPASTG 465
QY 318 -----PAMEQASSGGYLTW-----RHNGT-----TFPAGSVSVLPEGF- 351
DB 466 PTVSTADVTSPPTAGTSCASPVTPSPRDNGTESKAPDMSPISAVTTPPNQNSPPT 525
QY 352 ALERIDPNDG-----WTPDASAGDVTFRQYAVDEVYVYTNPNAG----- 391
DB 526 AMTTPPNATSPGLKTSPTSAVTTPPNATSPPTAVTTPPNATSPPTVGETSPQANATN 585
QY 392 ---GSAFTYRVPPSNAYTNTVFRNTLETRPSSRRLEMPAPDFCQYANPNKIQ 448
DB 586 HILGGSFPPVYTPPKNATSDV---TTGQHNKRTSSSTSSKLRSPSSIPETTSHPMLTS 642
QY 449 SILKETLGCYLVHSKMRNFVQLTPAS--SFGAVSFNNP 485
DB 643 A-----HPTGENITGVTPASISTHHVSTSP 669

RESULT 8

Q9RY11
ID Q9RY11 PRELIMINARY: PRT: 597 AA.
AC Q9RY11:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL 60.6 KDA PROTEIN.
GN DR0143.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.U., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE001876; AAF09732.1;
DR TIGR: DR0143;
DR InterPro: IPR001917; AminoTransf_2.
DR InterPro: IPR001344; DUF11.
DR Pfam: PF01345; DUF11.
DR ProSite: PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 597 AA; 60604 MW; B9A61442BCBA8279 CRC64;

Query Match 3.9%; Score 130; DB 16; Length 597;
Best Local Similarity 20.6%; Pred. No. 0.76;
Matches 128; Conservative 67; Mismatches 253; Indels 172; Gaps 25;

QY 54 ODLAQSLDANTVTFPANISSMPEFR-----NMAKKRIDSDSIGMYEKYLDPAAGT----- 105
DB 4 RDIYDLPAPAGGGWVSTSPYQAVCVNNSGGILISNNSYLSYSSDSQSGVPPA 63
QY 106 ----ESARAVGEYSKIPDGLVFSVDAIRE--IYNECP--VYVDVSVPLDGRWSISI 157
DB 64 PTAGTQMSVPNTIIPAGTCRFSFDVTLPRAGTYKKNYIPAIRISGGNDAGNASTTA 123
QY 158 FSPMERFAY-----VAVANVENKEMSLD--VNDLIMLNLMADRYV 200
DB 124 FAGPTVTKAYSPKSLIDGKATSTLTITLNNSSPAVASITAPLATDNIG--NGLETGYTT 181
QY 201 DSEQWINTNDTYYVIRIVLRP-----YVDVDPTEGLVTVSDYRLTYKAITCEANMP 255
DB 182 SCPTATPESGTTIYTPSGATLNPCTTATVRSATAG-----SY 222
QY 256 TLVDG---FWIGGOYA-----LTPSLPDYVSEAYA-----LHTLFARPSA 297
DB 223 NTKISGALQTYGNNRAAASDLTYTSTFRLLTITTHASQNTAGOTGYTYTVSNASGA 282
QY 298 AALAFVWAGLPQGTAPAGTAPAMEQASSGGYLTWHNGTTFPAGSVSVLPEGALERD 357
DB 283 AA-----TSGALSLDLPSCGMSFNSVT 305
QY 358 PN-DGSMWTFASAGDVTFRQYAVDEVYVYTNPNAGGASPTTVVPPSNATNTVFNKT 416
DB 306 TTAGSGFGRPPASGAT-----GRVDWTFPTSPPLAAGSLTFETVNVANVANATATLTN 360
QY 417 LLETSPSSRRLEMPAPAD--FGQYANPNKIEOSILKETLGCYLVHSKMRNFVQLT--- 472
DB 361 VASVGGGDPDVLPTFRGATCTGEOCASDPTV-----KRITQTLIRK 402
QY 473 --PASSFGAVSFNNPGEYERTRDLPDYTGIRDSFDQNMST-AVAHFRSLSHSCSIVTKTYQ 529

```
Db 403 EFGAGAPKGFN-----YDVATI--TISQGSSTLTATNTSSIANPGVATDT-- 449
Qy 530 GMEGVTVN-TPEGQFAGLKNKEIICLADLATRLTGVPATDNFAAFAANM- 587
Db 450 -----INIRPGANYTLREVLKNDVAFSGPDSYDSR-----YTCTNMTGTSTTWPNTSS 498
Qy 588 -LSSVLKSEATSSIIKSVGE 606
Db 499 GMSFTLTPOAGDIITCSVD 518

RESULT 9
ID Q95S10 PRELIMINARY; PRT; 466 AA.
AC Q95S10;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE H102328P.
GN DSCAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
EX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY061009; AAL28557.1;
DR SEQUENCE 466 AA; 50943 MW; 7056AB9B955CC2B CRC64;
SQ

Query Match 3.88; Score 127; DB 5; Length 466;
Best Local Similarity 20.7%; Pred. No. 0.84;
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;

Qy 231 TEGIVRTVSRYRLTYKAITCEANMPTLVDOGFMT-----GGQVALTPSLPQDYDSEA 283
Db 71 TETKTQKVPHYQMSFEATELEKNRP---YEFWTASTTIGEGQSKSIYAMPDQYPAK 126
Qy 284 YA-----LHTLTFARPSAALAEVWAGLPDGTAPAGTAPMEQASSGGYLTWRHNGTTP 339
Db 127 IASFDFTATFKEDAKMPCLA---VGAPQ-----PEITWIKGVERS 166
Qy 340 AGSVSYVLPGEFAL-----ERYDPNDS-WTDFASAGDTYTFROYAV----- 380
Db 167 ANDRMRYLPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIVLAPQSPHVTLSATP 226
Qy 381 -DEVVVTNNPAGGSAPT--FTVRVPPS-----NAYTN 410
Db 227 TDALTYKALKHEGDPALHGYTLHYKREFGEWETSEVSDQKNITGLGSRGYVA- 285
Qy 411 TVERN-----TLETPRSSRLLEMPRPFQGTVANPNKIEOSLKEITLGC---YL 459
Db 286 TGFNNIGAGBASDILNTRTGQKPLPEKPR-FIEVSSNSVSLHFKWMKG-GCPMSHFV 343
Qy 460 VHSKMRNP-----VFQLTAPASSFG--AVSFNNPGY----- 487
Db 344 VESKRRDQIEWNOISNNVKNVYVLDLEPATWYMLRITAHNSAGFTVAEYDFATLTVT 403
Qy 488 -----ERTRLPD 495
Db 404 GGTIAPSRLPE 415

RESULT 10
Q95TGO 10
```

```
ID Q95TGO PRELIMINARY; PRT; 1212 AA.
AC Q95TGO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE S09407P.
GN DSCAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
EX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY058789; AAL14018.1;
DR SEQUENCE 1212 AA; 134066 MW; 6D9061E62EF9B5C CRC64;
SQ

Query Match 3.88; Score 127; DB 5; Length 1212;
Best Local Similarity 20.7%; Pred. No. 3.6;
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;

Qy 231 TEGIVRTVSRYRLTYKAITCEANMPTLVDOGFMT-----GGQVALTPSLPQDYDSEA 283
Db 455 TETKTQKVPHYQMSFEATELEKNRP---YEFWTASTTIGEGQSKSIYAMPDQYPAK 510
Qy 284 YA-----LHTLTFARPSAALAEVWAGLPDGTAPAGTAPMEQASSGGYLTWRHNGTTP 339
Db 511 IASFDFTATFKEDAKMPCLA---VGAPQ-----PEITWIKGVERS 550
Qy 340 AGSVSYVLPGEFAL-----ERYDPNDS-WTDFASAGDTYTFROYAV----- 380
Db 551 ANDRMRYLPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIVLAPQSPHVTLSATP 610
Qy 381 -DEVVVTNNPAGGSAPT--FTVRVPPS-----NAYTN 410
Db 611 TDALTYKALKHEGDPALHGYTLHYKREFGEWETSEVSDQKNITGLGSRGYVA- 669
Qy 411 TVERN-----TLETPRSSRLLEMPRPFQGTVANPNKIEOSLKEITLGC---YL 459
Db 670 TGFNNIGAGBASDILNTRTGQKPLPEKPR-FIEVSSNSVSLHFKWMKG-GCPMSHFV 727
Qy 460 VHSKMRNP-----VFQLTAPASSFG--AVSFNNPGY----- 487
Db 728 VESKRRDQIEWNOISNNVKNVYVLDLEPATWYMLRITAHNSAGFTVAEYDFATLTVT 787
Qy 488 -----ERTRLPD 495
Db 788 GGTIAPSRLPE 799

RESULT 11
Q9V4J9 PRELIMINARY; PRT; 2016 AA.
AC Q9V4J9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CG17800 PROTEIN.
GN DSCAM OR CG17800.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
EX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
```

[illegible]

		3.88;	Score 127;	DB 5;	Length 2016;
		Best Local Similarity	20.7%;	Pred. No. 7.9;	
	Matches	77;	Conservative	45;	Mismatches 116; Indels 134; Gaps 18;
OY	231	TEGLIVRTSDRLTYKATITCAENPTLVDOGFMT-----GGQYALPTPSLPDYDSEA	283		
Db	1259	TETKTQKVPHYMOSEATELEKNRP-----YEFWTASTTIGGGQOSIYAMPEDQYPAK	1314		
OY	284	YA-----LHTLTFARSSAALAFYWAGLPQGGTNPAGTPRAMEQASSGGYLWLRNGTTFP	339		
Db	1315	IASEDDTFTAFKEDAKMPCLA-----VGAPQ-----PETWIKIKGVFEFS	1354		
OY	340	AGSVSYVLPEGFAL-----ERYDPNDGS-WMDFASAGLTVFEROYAV-----	380		

Dd	1355	ANDRMKVLDEGDSLLIKSVNRQDAGVYSCHAENSIADSKITHKIIVLAPQSPHVTLSATP	1414
Oy	381	-DEVYTNPNPAGGSGA-PT--FTVRVPSS-----NAYTN	410
Dd	1415	TALTVYKLPHGEDTAPLHGTYLHYRPEFGEMETSEVSVDOSKNHTIEGLCGRYOYA-	1473
Oy	411	TVFRN-----TLLETSPSRRLLEPMPPADGGOTVANNPKIEQSILKETLGCC---YL	459
Dd	1474	TEFNNGIGAEASDIINTRRKGCKPKLPEKPR-FIEVSSNSVSLHFRAWMDG-GCPMSHY	1531
Oy	460	VHSKMNP-----VFQLTPASSFC--AVSENPGY-----	487
Dd	1532	VESKRQDJEMNQISNNVXPDNNVVYLDLEPATWYMLRTAHNSAGFTVAEYDFATLTWT	1591
Oy	488	----ERTDLDP	495
Dd	1592	GCTIAPSRLDPE	1603
 -RESULT_13			
AC	099552	PRELIMINARY:	PRT: 3570 AA.
DT	01-MAY-1997	(TREMBLrel. 03, Created)	
DT	01-MAY-1997	(TREMBLrel. 03, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	MUCIN MUC5B (FRAGMENT).		
GN	MUC5B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PLACENTA;		
RX	MEDLINE=97166151. Pubmed=9013550;		
RA	Desseyn J.L., Guyonnet-Duperat V., Porchet N., Aubert J.P., Lalre A.;		
RT	"Human mucin gene MUC5B, the 10.7 kb large central exon encodes		
RT	various alternate subdomains resulting in a super-repeat. Structural		
RT	evidence for a 1lp15.5 gene family."		
RL	J. Biol. Chem. 272:3168-3178(1997).		
DR	EMBL: 272496; CNA96577.1; .		
FT	NON_TER	1	
FT	NON_TER	3570	
SQ	SEQUENCE	3570 AA; 361019 MW; DE04E4D727579312 CRC64;	
 Query Match 3.7%; Score 126; DB 4; Length 3570; Best Local Similarity 22.3%; Pred. No. 22; Mismatches 178; Indels 118; Gaps 14 Matches 94; Conservative 32;			
Oy	221	LRPITVDPPPTGGLAKRVSYDYRLTY-----KAITCEANMPITLVDOGFWGIGQYALTPTSIP	276
Dd	643	LRSTATTTPTATSFIAIPSSSLGTMTWRLSQTTPPATMST-----ATPSSIRP	689
Oy	277	QYDVSEANAHLTL-----FARSSAAALAFVNAGLPDGGTAPAGTPAMEQ	322
Dd	690	E-----TVHSTVLTATTATTGATGSAVTPSSPTGTAHTTKVLTITTTTGCFATPSSSP	742
Oy	323	ASSGGYLMRNHNGCTFPAGSVSYVLPEGFALRIYDPNDOSMWDFASAGDTVFFROVAUDE	382
Dd	743	GHARTLPWA-IISTTPTPTTRGSTVPSST-----PGTHHPVYLTITTT	784
Oy	383	VVVTNNPAGGSGAPFTVVRVPSPSNVYTNVFNNTLLETSPSSRRLLELPPPADFGQTVAN	442
Dd	785	TTVAVGSMATPSSSQTSGTSPSLTTTATTTATIGSTTPSSPTGGTTPPV--LTTTAT	842
Oy	443	NPKIEQSILKETLGCYLVHSKMRNPVFLTPASSFGAVSFNNPGYERTFDLPDYTGIRDS	502
Dd	843	TPAATSS-----TVTPSSALGTT-----HTPPVPN-----	867
Oy	503	FDOANNSTVAHRSISHS-----CSIVTKTYGWMGVNVNTPFQGFAN-----AGLKN	552

```

Db      868  -----TLATTHGRSLSPSSPHVTCAWTATSGLIGTHTHEPSTGTSHTPATGTGTH 922
QY      553  EETILCLADDLATRLGLVPA-----IDNFAANVSAPAA-----NMUSVLKSEATSSITK 602
Db      923  STPALSPHPSRTTESPPSPGTTTGGHTTASRTTATATPSPKTRTSTPOPTAPIT 982
QY      603  SV 604
Db      983  TV 984

RESULT 14
Q9KHB3 PRELIMINARY; PRT; 717 AA.
AC Q9KHB3:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CHITINASE.
GN CH11.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-2;
RA Wang Y.L., Wang H.Y., Qin M., Zhang Y.Z.;
RT "Nucleotide sequence of a chitinase gene (chl1) from Bacillus
RL submitted (May-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR EMBL; AF255220; AAF4782.1; -.
DR HSSP; P02751; 1TTF.
DR InterPro; IPR001579; Chitinase_2.
DR InterPro; IPR003610; Chitin_bind3.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF02839; CBD_7; 1.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00495; ChEBD3; 1.
DR SMART; SM00600; FN3; 2.
DR PROSITE; PS01095; CHITINASE_18; UNKNOWN_1.
KW Hydrolyase; Repeat.
SQ
SEQUENCE 717 AA; 76118 MW; BB0B019CDE72C198 CRC64;

Query Match 3.7%; Score 125.5; DB 2; Length 717;
Best Local Similarity 21.6%; Pred. No. 2.1;
Matches 117; Conservative 61; Mismatches 188; Indels 175; Gaps 26;
QY 195 DMRYVV-----DSBQWNTFNDTYYVIRIRLRTYGVPPTEGLVTVSDYRLTYRAI 248
Db 202 NMEYVSGGLDGNNSRPPDKQNYTLTLKIRKRLDAEAIVGKEYLLTLIASASPTYAAN 261
QY 249 TCEANMPLVDOGVITGQIALTPSLDPQDYSEAY---ALHTLTFARPSSAALAFVWA 305
Db 262 TELNIASTIV--WI-----NIMTYDFNGAMOKISAHNAPLNDPAAPAS-----A 304
QY 306 GLPQGGT-----APAGTPA-----WE-----QASSGGYLTMRHNGT--T 337
Db 305 GVPDSNFTFNVAAGAGHILNAGYPAKKLVLPFFYGRGWDGCAQANNNGOYOTSGGSSIGT 364
QY 338 FPAGS-----VSYLPEGFALERYDPNDGSWTD-----FASAGDT----- 372
Db 365 WEAGSFDYDELANVINKNGYT--RY-----WMDTAKVPLLYNASNKRFLISYDDAESIG 416
QY 373 -----VTFROYAVDEVVYVYVNNP-----AGGGSAPTFYVRVP--PSNATYNTV 412
Db 417 HKTAYIKSKGLGAMFWELSDRKNKTLKLSDSLSTGSGVPPPTDTTAPSPVGNARSTGV 476

```



```

QY 413 FRNTL-LETRPSSRLLELPPADFGQFVANNPKIEQSLKRETL-----GCYLVRK 463
DB 477 TASSVTIAMNASTDNVGV-----GYTYNGTSLVSTGTATITISGLAPGTSTFTYK 530
QY 464 MENPVQLTPASSFGAVSFNNPGYERTDLPDYGIROSEFQNMSTAVAHFRLSHSCSI 523
DB 531 AKDAGNLSAASNSLTVS-----TTVQPGDYGQAPTVPTNLSTFAKTSST 575
QY 524 VTKT-----YQMGVTVNTPFGQFAH-AGLKRHELLCLADLATRLTGYY 570
DB 576 ILLSMAASTDNVGVIGYEYNGTALVTYVSGTSATVTL-----TADTSYTFVAK 627
QY 571 PATDNFAAASFAANMNLSEVLSKSEATSSIIKSGET--AVGAQSGLAKLPGLMSVYG 628
DB 628 DAAGNLASAASSAL-----TYKTE-----VGTINPGVSAKMANATAYVQGLVTVNG 672
QY 629 K 629
DB 673 K 673

```

RESULT 15

```

Q97206 PRELIMINARY; PRT; 1269 AA.
AC Q97206;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE PROTEASE RELATED PROTEIN.
GN SS01141.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Weiher C.C.Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Th-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL: AE006730; AAK1392.1;...
DR InterPro: IPR000209; Peptidase_S8.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1269 AA; 138027 MW; 87C8F05CFCF83DA84 CRC64;

```

Query Match 3.7%, Score 125.5; DB 17; Length 1269;
 Best Local Similarity 19.6%, Pred. NO. 5;
 Matches 98; Conservative 69; Mismatches 181; Indels 153; Gaps 27;

```

QY 18 RNVRSATVTVNGRRNRRTGRGVSPDNFTAAQDLAOSLDANTYTFPANISSMPEF 77
DB 584 QRLVTAITITPNC-----SPVQ--TGEFKALIEYLGNTLTFTNTYNSLTKL 629
QY 78 RNMAGKIDLDSDSIGWFKYLLDPAGATESARAVEYSKIPDGLVFKFSVDAREIRY--- 134
DB 630 --WT-GSGVLNKAAGLFFVYV--YGSSDGLRGIGYTFEGSYITFTFVYVEL 684
QY 135 -NEECPVTVDS---VPLDGQWMSLIFSFPMEFRTAIVAVAN---VENKESLDVNDL 186
DB 685 GNAELGITLSNSYFQAPIGVANNITLNTSYNTTNAFYFTLLSVIRKNGVGDIDLPDL 744
QY 187 IEWLNLADMRVYVSEQW--INFNDT--TYVVRIV-----LRP 223
DB 745 -----SIDDLLIEGNAKYGDAFTNGVYMOTLFIIPVYVPEGSVSPGHITTEGSIIP 799

```

```

QY 224 TYVDVDPPT--EGLVRTVSQDYRLTYKATNC-----EAMNPTLYDQGFNIGGOYALTPRSL 275
DB 800 PVNLPSSTTFODALOGT-----NITAKLVSSNGGVINEANIP-LSPNGIYFG--YLXIPKNT 852
QY 276 PQ-YDV-----SEAYALHTLTFARPSSAAL-----AFVWAGLPQ 309
DB 853 PSGLYNVLLFATYYSYTLNTTIRGFGYQIYVSNGATISVKSVMYAFEGQVFIYANTIN 912
QY 310 CGTA-----PAG-----TPAEQASSGGYLTWRHNGT--TFPAGS 342
DB 913 GTNEIKFGMFSAFVYVPSLSLSEFNYYTIIIEIPLWNPKIG--EMEGNFTLPSAISAGN 969
QY 343 VSYVLEPGFALERYDPNNGSWTDFASAGDTYFFRQVAADEVVYVNTNPPAGSGSAPFEYRV 402
DB 970 LTYLAGOGTF-----GVPER-VLITGISALGNPTTNSGNAYTTINV 1009
QY 403 PPSNAYTNTVFPRNTLLETRPS 423
DB 1010 LPTLTFTN-----QLDKTLPS 1026

```

Search completed: May 29, 2002, 11:41:27
 Job time: 159 sec

